Università di Genova

UNIVERISITY OF GENOA

DOCTORAL THESIS

Genomic analysis of special types of invasive breast cancer

Author:

Lorenzo FERRANDO

Supervisor:

Dr. Gabriele ZOPPOLI

The Translational Genomic Lab

Department of Internal Medicine

20-12-2021

Declaration of Authorship

I, Lorenzo FERRANDO, declare that this thesis titled, "Genomic analysis of special types of invasive breast cancer" and the work presented in it are my own. I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Abstract

Breast cancer (BC) is a heterogeneous disease encompassing a wide array of categories of cell subpopulations, molecular aberrations and different clinical and pathological behaviors. The most challenging obstacle faced by clinicians is represented by a full understanding of such heterogeneity and processes of resistance induced after the administration of anticancer drugs. For such reasons, BC is classified according to a number of several clinico-pathological and molecular factors. Over 20 histologic special types of BC are recognized by The World Health Organization (WHO) [1].Special types of BC account for approximately 20% of all newly diagnosed cases. In spite of these mostly morphohistological classifications, treatment decisions are mostly guided by the size of the primary neoplasm at presentation, the involvement of locoregional lymph nodes, the status of few, actionable biomarkers such as estrogen and progesterone receptors and the amplification of the *ERBB2*/HER2 gene, as well as by the degree of differentiation and the proliferation index of cancer cells. The morphology of BC cells and their histological architecture, accounting for the identification of a special type, is almost not taken into account in clinical decisions by most guidelines [2]In this thesis, I focused on the molecular characterization of special types of BC, with the purpose of identifying potential novel biomarkers of diagnosis, prognosis, and therapeutical prediction in such large, but in part neglected, class of tumors. Out of my research, I published two original scientific papers. The first one describes the genomic landscape of four special types of BC: invasive lobular carcinoma (ILC), mucinous carcinoma (MUC), micropapillary carcinoma (MPC) and metaplastic carcinoma (META). The second one focuses on the molecular comparison between metaplastic carcinoma and uterine carcinoma, stemming from their similar histology and clinical behavior.

Acknowledgements

I will always be grateful to my mentor and friend Gabriele, for his support and friendship.

I want to thank David, my great friend, for his guidance and advice.

I would like to say thanks to Giorgia, who always inspired me for her determination and dedication.

I want to express my infinite gratitude to my parents, without whom this journey would not have been possible.

Contents

List of Figures

List of Abbreviations

BC	**B**reast **C**ancer
ER	**E**strogen **R**eceptor
PR	**P**rogesterone **R**eceptor
HER2	**H**uman **E**pidermal growth factor **R**eceptor 2
HR	**H**ormone **R**eceptor
mBC	**m**etastatic **B**reast **C**acer
ILC	**I**nvasive **L**obular **C**arcinoma
MUC	**MU**cinous **C**arcinoma
MPC	**M**icro**P**apillary **C**arcinoma
META	**META**plastic carcinoma
IDC-NST	**I**nvasive **D**uctal **C**arcinoma **N**on **S**special **T**ype
UCS	**U**terine **C**arcino**S**arcoma
TNBC	**T**riple **N**egative **B**reast **C**ancer
ET	**E**ndocrine **T**herapy
AI	**A**romatase **I**nhibitors
SERD	**S**elective **E**strogen **R**eceptor **D**egraders
HRD	**H**ologous **R**ecombination **D**eficiency
MPS	**M**assively **P**arallel **S**equencing
TSG	**T**umor **S**uppressor **G**enes
TMB	**T**umor **M**utation **B**urden
FFPE	**F**ormalin **F**ixed **P**araffin **E**mbedded

Chapter 1

Introduction

Breast cancer (BC) is a heterogeneous disease encompassing a wide set of categories of cell subpopulations, molecular aberrations and different clinical and pathological behaviours. Thereforse, BC is classified according to a number of several clinico-pathological and molecular factors. In this section, molecular and histologic characterization is explored.

1.1 Molecular characterization of breast cancer

In 2000, the pioneering work of Perou et al. [5] and Sørlie et al. [6] demonstrated that BC can be subdivided into at least four intrinsic molecular subtypes, that reflect differences in phenotype, prognosis and treatment response. These seminal studies analyzed the gene expression differences through cDNA microarray approach. Based on the results, the Authors discovered that their classification was mostly driven by the expression of three membrane receptors, namely the estrogen receptor (ER), the progesterone receptor (PR) and the epidermal growth factor receptor 2 (*ERBB2*/HER2), as well as by the regulation of cell proliferation (summarized by the proliferation-related protein Ki67). Collectively, ER and PR are considered the main hormone receptors (HRs) in BC. These elements represent one of the pillars for BC classification as of today. In clinical practice, the activity of HR/HER2/Ki67 is evaluated by a combination of

immunohistochemistry (IHC) and in-situ hybridization (ISH), which are considered sufficient and reasonably robust to frame the clinically relevant molecular characteristics of BC and drive therapeutic choices. Nevertheless, due to the increasing awareness of the huge interpatient variability of BC, the demand for more refined and specific classification of BC has led to the development of the so-called multigenic assays such as: PAM50 (50 genes), Mammaprint (70 genes) and Oncotype DX (21 genes) [7-9], allowing for the simultaneous evaluation of the status of dozens of genes. These multigenic assays were designed to identify, with the technologies available in the first years of the 2000s, early stage BC cases at higher risk of recurrence and thus deriving greater benefit from adjuvant chemotherapy. The four intrinsic molecular subtypes identified by Perou et al. [5], and consistently showing biological and clinical coherence over the last twenty years are: luminal A, luminal B, Basal-like, HER2-enriched and normal-like.

- **LUMINAL A** is the most common intrinsic subtype [10]. According to the current guidelines, luminal A is characterized by the expression of ER ($\geq$ 1%) and/or PR ($\geq$ 20%), poor expression of HER2 ($\leq$ 30%) and low levels of Ki67 [11]. Generally, Luminal A cancers are low-grade, and are associated to good prognosis, slow growth and limited lymph node involvement [12]. The positive status of HR makes luminal A targetable by endocrine therapies (ETs) such as aromatase inhibitors (AI), or selective estrogen receptor degraders (SERDs) [13].

- **LUMINAL B** is responsible for 20-30% of invasive BC [10]. This subtype is characterized by ER ($\geq$1 %), PR (<20% or any) and HER2 ($\leq$ 10%) and high level of Ki67 ($\geq$ 20%, although IHC cutoffs are varied and there is no universal consensus) [11]. Compared to luminal A, luminal B BC shows a higher growth rate and worse prognosis with higher chance of local and distant recurrence [12]. It is considered

the most aggressive form of HR+ BC and does not always derive benefit from ETs, although in the absence of biomarkers of primary resistance, ETs are offered to all patients falling in this category [14]. The main molecular differences between the two luminal types is represented by the increased activity of genes involved in the proliferation process (*NSEP1*, *CCNE1*) and activation of PI3K pathway in luminal B cancers [15].

- **BASAL-LIKE** BC constitutes from 10 to 20% of BC [10]. It is characterized by the lack or very low levels of expression of ER, PR and HER2 ($< 10\%$), but high proliferation indices (ki67 $> 30\%$ with few exceptions, such as apocrine BC) [11]. For this reason, they can be defined as triple-negative BC (TNBC), although there is no complete overlap between TNBC (which is an immunohistochemical definition of BC) and basal-like BC (which is deemed so by gene expression profiling). Basal-like cancers are aggressive and show a high risk of both local and distant recurrence, especially in the first few years after diagnosis [16]. Their diagnosis subtends a generally poor prognosis. More recent studies have focused on the molecular profiling of this category of BC, further identifying several subclasses: basal-like 1 and basal like-2 (BL1 and BL2) display TNBC gene expression patterns but a different immune activity; luminal androgen receptor-like (LAR) is characterized by specific a specific gene expression profile of androgen metabolism; immune-modulatory class (IM) has altered activity of immune related signatures; two subtypes are associated to an enrichment of the EMT (epithelial mesenchymal transition) pathway, and are called mesenchymal (M) and mesenchymal stem-like (MSL) [17]. The stability of these further classifications has been rediscussed recently, but its overall validity holds true [18]. In the clinical practice, TNBC/basal-like BC represents a real challenge

for its molecular and morphological heterogeneity and the current lack of targeted therapies [19].

- **HER2-ENRICHED** BC is responsible for 15% to 20% of cases [10]. Positivity of HER2 is usually evaluated: 1) in IHC as strong/complete membrane staining (score 3+) in at least 10% of cancer cells or 2) in ISH, where the HER2/CEP17 ratio ($\geq$ 2) and an average *ERBB2* gene copy number ($\geq$ 4) determine the amplification of HER2 [20]. In clinical practice HER2 status is assessed by IHC and ISH in case of intermediate protein staining. HER2-enriched cancers are associated to a high proliferation phenotype and have a poor natural history. However, HER2+ cancers can be successfully treated with targeted therapies focusing on the HER2 protein complex (Trastuzumab, Pertuzumab, Lapatinib, Neratinib) [19], which have largely reversed such prognosis in clinical practice over the last twenty years [21].

1.2 Histological types of breast cancer

BC can be classified according to the tissue morphology. The main determinant of invasiveness consists in defining whether the malignancy is circumscribed to the epithelial part of the breast (so-called in situ BC, which is a misnomer for a premalignant lesion), or extents to the stroma, and if the tumor invades ducts or lobes [22]. However, in clinical practice, rather than the precise location, several factors such as cell type morphology, cancer architecture, secretion activity and IHC define the histological nature of the tumor. The most common histological form of BC is the invasive ductal carcinoma (IDCs). IDCs can be classified as 'not otherwise specified' (IDS-NOS) or 'no special type' (IDC-NST), due to the absence of peculiar morphological characteristics. In fact, IDC-NST is a heterogeneous disease entity, and shows a wide range of clinical behaviors. In contrast,

IDCs can be classified as of pertaining to a 'special type', whenever the cancer tissue presents distinctive characteristics. Histologic special types of BC account for up to 20% of all BC cases. The most common forms are: invasive lobular carcinoma (ILC), metaplastic carcinoma (META), mucinous carcinoma (MUC), micropapillary carcinoma (MPC).

- **INVASIVE LOBULAR CARCINOMA (ILC)** is the second most common of distinct BC, representing approximately 10-15% of newly diagnosed BC [23]. It is characterized by the infiltration of discohesive cordon-structured neoplastic cell in the lobules wall and in the mammal tissue and by the loss of E-caderin (*CDH1*) staining in 95% of cases. ILC is more common in older woman, tending to occur later in life than IDC [24], is less sensitive to chemotherapy, may respond better to Ais, and has a tendency to relapse later than IDC and in unusual sites, such as submucosal tissues, uterus and eye socket [25].

- **METAPLASTIC CARCINOMA (META)** is characterized by a predominant component of metaplastic differentiation. It is a poorly differentiated tumor, showing heterogenous morphological features, containing ductal component mixed with squamous/spindle cells or chondroid cells. Often it is associated with lymph node involvement and has an aggressive behavior. It comprises about 1% of BC, occurring most likely after menopause [24].

- **MUCINOUS CARCINOMA (MUC)** , also known as mucoid carcinoma, represents about 4% of BC. It has been associated to a better prognosis. MuC is characterized by abundancy of extracellular mucins that cluster around tumor cells [24].

- **MICROPAPILLARY CARCINOMA (MPC)** represents about 6% of BC [24].

It is one of the most aggressive special types and is associated to unfavorable prognosis due to the extremely early and frequent involvement of axillary lymph nodes [26].

Chapter 2

Next generation sequencing: an overview

In the last decade, the tremendous advances of massively parallel sequencing (MPS) technologies have allowed for the analysis of large case sets with a significant decrease in costs, an improvement of turnaround times, and an increase in sequencing accuracy. Nowadays, oncologists rely on MPS for clinical purposes as well as for research. In this chapter, I describe the most common aberrations of BC with potential or practical clinical consequences.

2.1 Oncogene and tumor suppressor genes

Cancer-related genes can be divided into two classes according to the different role played in the biology underlying a specific cancer type. The first class is that of proto-oncogenes. This class consists of genes coding for proteins involved in cell cycle regulation, cellular differentiation, and cell death. The transition from proto-oncogene to oncogene may lead to the neoplastic phenotype, especially when several oncogenes act at the same time. Bona fide examples of oncogenes in BC are *ErbB2*, *MYC* and *PIK3CA*. The transition of a proto-oncogene to a cancer driver gene occurs upon even slight modifications of its wild-type form such as: 1) Mutations

within the body of the proto-oncogene or within its regulatory region: the mutation may lead to enhancement of the coded protein activity or to an abnormal active regulation. 2) Altered number of genetic copies of the proto-oncogene, with a resultant overproduction of its protein product. 2) Increment of the concentration of a protein, caused by altered gene expression levels, or abnormalities in the protein stability, causing a prolonged activity in the cell. 3) Genome rearrangement such as fusion and translocation. The second class of cancer genes is that of tumor-suppressor genes (TSG). A TSG codes for a protein involved in fundamental biological processes during cell division (e.g. repairing DNA error) or apoptosis. When a TSG is mutated, it results in its inactivation, leading to the development of neoplasia. Contrary to oncogenes that are characterized by a 'gain of function' alteration, TSGs are affected by a reduction or a 'loss of function'. Another fundamental difference between oncogenes and TSGs is summarized by the so-called Knudson's double-hit hypothesis. According to this hypothesis, both alleles of the same gene must be affected before an effect is manifested, while only one gene must be activated before it manifests its obnoxious. In other words, as long as one of the alleles of the gene is not damaged, its function is preserved. The most known TSGs in BC are *BRCA1*, *BRCA2*, *PTEN*, *TP53*.

2.2 Gene Mutations

DNA sequencing has allowed to determine the entirety of an organism genome. In other words, it is possible to know the exact status (mutated or wild-type) of each gene of an individual at a given time. However, in clinical practice such gigantic amount of information is limited to a handful pool of genes with an established association with cancer disease. For

this reason, the last decade has seen the rise of the so-called targeted sequencing. Targeted sequencing consists in analyzing specific genomic regions known or suspected to play a putative role underlying a specific phenotype or a disease. In BC, approximately 50 genes are considered meaningful for clinical purposes. This list of genes comprises both oncogenes and TSGs such as *BRCA1/2*, *TP53*, *PIK3CA*, *ESR1*, *PTEN*, *PALB2*, *GATA3*, *KMT2C*, *NCOR1*, *AKT1*, *NF1*, *CDH1*, and *RB1*. In recent studies, a number of novel genes with likely oncogene or TSG roles, including *TBX3*, *RUNX1*, *CBFB*, *AFF2*, *PIK3R1*, *PTPN22*, *PTPRD*, *SF3B1*, and *CCND3*, have been reported [27, 28]. One of the most known TSG in BC are the *BRCA* genes. *BRCA1/2* play a key role in the regulation of the cell cycle, cell proliferation, cell differentiation, and, importantly, in the process of DNA damage repair (DDR) by homologous recombination. DDR maintains nucleic acid sequence accuracy during DNA replication and transcription, as well as upon cell exposure to potentially harmful external agents [29]. Loss of function of BRCA genes results in homologous recombination repair deficiency (HRD), one of the most well-known molecular mechanisms underlying cancer development and progression. Another recurrent mutated gene in cancers is *TP53*. Mutations affecting *TP53* lead to a loss of function effect [30], and in the context of BC, are frequent in estrogen receptor (ER)-negative breast cancer. One of the most frequently mutated oncogenes in HR+ BC is the Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha, *PIK3CA*. *PIK3CA* codes for the p110α subunit of PI3K. *PIK3CA* mutations within the helical and the kinase domains cause a positive activation of PI3K pathway, resulting in uncontrollable cell proliferation [31]. Similar biological effects result from *PTEN* loss-of-function [32], as this tumor suppressor protein is a negative regulator of *PIK3CA*. Another clinically relevant and frequently mutated gene in BC is *ESR1*. Mutations affecting *ESR1* cause a constitutive activation of the Estrogen Receptor alpha (ERα), which results in enhanced

cell proliferation and resistance to endocrine therapy in ER-positive breast cancer [33]. Genomic analysis highlighted that *ESR1* mutations are found almost exclusively in metastatic HR+ BC rather than in primary tumors and have paved the way for the evaluation of *ESR1* mutational status for its potential clinical utility [34,35]. *ESR1* mutations are of special interest, because they are a proof of principle of the evolutionary mechanism of cancer escape under treatment pressure.

2.3 Gene Amplifications/Deletions

Copy number variation (CNV) represents a type of structural variation in which genomic sections are repeated altering the hypothetical physiological number of alleles of a gene. CNV encompasses two types of modification, amplification and deletion. In BC, CNVs represent one of the most predominant genomic modifications of clinical relevance [36]. An example of a CNV of prognostic and therapeutic value is the amplification of the HER2 (*ERBB2*) gene [37], located on the chromosome 17 (q12). *ERBB2* amplification causes the overexpression of the tyrosine kinase receptor HER2, which acts as a positive trigger for the downstream activation of PI3K/AKT and MAPK pathways. These, in turn, play a crucial role in tumor cell proliferation and survival. Patients with HER2-enriched BC benefit from the use of anti-HER2 monoclonal antibodies. The standard assessment of CNV is based on quantitative PCR or ISH, but such techniques only allow to detect specific DNA sequences. Multi-target CNV evaluation can be performed using comparative genomic hybridization (CGH) technique and whole genome single nucleotides polymorphism (SNP) array, or alternatively, using NGS approaches. Besides ERBB2, other potentially actionable genes often amplified in BC are *NOTCH1/2/3*, *MYC*, *FGFR1/2*, and EGFR. The most common deletions found in BC involve *PTEN*, *CDH1*, *CDKN2A/2B*, *RUNX1/CBFB*, *RB1*, *TP53*, and *INPP4B* [27].

2.4 Genomic Rearrangements

Another type of structural modification is the rearrangement of the genome. Genomic rearrangement results from translocation of DNA regions that lead to the formation of fusion genes. Although fusion genes have been mainly investigated in hematological malignancies, such as leukemias and lymphomas (e.g. *BCR-ABL1* fusion protein in chronic myeloid leukemia), they have also been detected in solid tumors, such as *EML4-ALK* fusion in lung adenocarcinomas [38]. Therefore, fusion gene detection is becoming a key aspect of genomic investigation in BC research. Genomic rearrangements can play an important role in cancer development and progression, and they may be clinically actionable genomic alterations with potential therapeutic impact. The standard assessment for the presence of fusion genes consists in FISH and IHC techniques, which might not be sufficient for multiplex analysis and more comprehensive tumor profiling. Nowadays, NGS has been implemented for detecting putative fusion genes and for finding potential new rearranged genomic regions [39]. In BC, few fusion genes have been detected so far [40]. Large genomic rearrangement involves for example *BRCA1/2*, *MAGI3-AKT3*, *FGFR3-TACC3*, *BCL2L14-ETV6*, and *ESR1-CCDC170*. For example, *MAGI3-AKT3* fusion results in the constitutive activation of AKT kinase, and it is now considered a possible target for ATP-competitive AKT small-molecule inhibitors [41]. Recurrent rearrangement found in ER+ BC is *ESR1-CCDC170* fusion (6–8%) of luminal Bs [42]. Currently, therapies able to target this specific type of chimeric protein product are not yet available and elucidative studies are expected to shed light on the various *ESR1* fusions, as well as their role in endocrine resistance of primary and metastatic ER+ BC. Finally, a family of rearrangements, namely the NTRK fusions, is believed to be an actionable genomic target in different cancer types including BC and has been the first genomic aberration to receive an agnostic approval of target drug,

i.e. its finding allows for selective inhibitors to be administered independently of the tissue of origin of the neoplasm [43]. However, its frequency in BC is very low.

2.5 Genomic signatures

A gene signature is defined as the combined action of a set of altered genes, or genes module, the effect of which modifies critical aspects of the cell. This process may occur as a result of abnormal biological process or pathogenic condition. An example of gene signature is the mutational genomic signature, that represents the effect of mutational processes on cancer due to the exposure to exogenous or endogenous mutagens. SNVs and chromosome CNVs are examples of mutational patterns that may affect genes involved in fundamental biological mechanisms. In cancer research, the mechanisms of DNA damage response, DNA repair, and DNA replication [44] are among the most relevant ones from a biological and clinical point of view. In BC, homologous recombination deficiency (HRD) plays a fundamental role [45]. HRD is strongly associated with the loss of function of *BRCA1/2*, *RAD51*, *PALB2* genes. HR is one of the main mechanisms of DNA repair after damage-caused double-strand breaks (DSBs). HRD detection is carried out on several different technologies such as array-based comparative genomic hybridization (aCGH) and NGS. HRD has gained the attention of the clinical and research community because it is more and more considered as a valuable biomarker for the response to DNA-damaging agents such as platinum agents and anthracyclines, as well as to PARP inhibitors [46]. Currently, two FDA-approved assays exist that allow to detect HRD along with other important signature, the telomeric allelic imbalance (TAI) [47], and large-scale transition (LST) [48]. Another critical signature is the tumor mutation burden (TMB), which is defined as the number of mutations per megabase (mut/mb) in a given

neoplastic specimen [49-50]. It has been already demonstrated that TMB acts as a clinically useful biomarker for immune checkpoint inhibitor (ICI) response, independently of PD-L1 expression, in NSCLC, melanoma, and bladder cancer [51,52]. This phenomenon can be traced back to the intrinsic nature of the TMB: tumors with high values of TMB are more likely to express new immunogenic antigens. This, in combination with ICIs, can lead to an activation of the immune system against cancer cells [53]. Although BC seems less immunogenic compared to other malignancies [54], it has been demonstrated that TMB can also be a predictive biomarker for ICI treatment and prognosis specifically in TNBC due to its high mutation rates. TMB assessment is often performed using dedicated gene panels like the Oncomine Tumor Mutation Load Assay (ThermoFisher®), SureSelect XT HS custom TMB and human all Exon v6 panel (Illumina®), QIAseq TMB panel, and NEOplus v2 RUO panel. Despite the limitations imposed by the analyzed circumscribed genomic regions, these panels were optimized to provide a reliable approximation of TMB. The rule of thumbs suggests that targeted panel size should be at least 1.1 Mb in order to provide an accurate estimate of the TMB [49]. Below the 1.1 Mb threshold panels overestimate TMB assessment, especially in those cancer types characterized by low or intermediate levels of TMB. A crucial aspect to consider during TMB estimation is the nature of the preservation applied to the tumor biopsy sample. Formalin-Fixed Paraffin-Embedded (FFPE) tissues represent a staple of research and therapeutics in terms of biospecimen preservation. Eventually, FFPE preservation leads to degradation of the DNA sequence in a process called deamination. Deamination introduces artefacts that alter TMB estimate. TMB is emerging as valuable biomarker for the stratification of ICI response. However, standardization in its assessment is still lacking in terms of harmonization and normalization across platforms and systems [55].

Chapter 3

The genomic landscape of metastatic special types of breast cancer

In this chapter, I focused on the study of four forms of special types of breast cancer. I studied their genomic profiles in both primary and metastatic forms and performed a comparison with a clinically matched cohort of IDC-NSTs. The study allowed a re-analysis from a specific perspective of the clinical cohort of the Razavi et al. study, available on cBioportal (https://www.cbioportal.org) [56,57,58].

3.1 Introduction

Breast cancer (BC) is heterogeneous and comprises various entities with divergent phenotype, biology, and clinical presentation [59,60]. There are over 20 histologic special types of BC recognized by The World Health Organization (WHO), accounting for 20% of all BCs [1]. Large sequencing studies have focused on invasive ductal carcinoma of no special type (IDC-NSTs), the most common histologic form of BC [28, 60-65], and data on the genomic landscape of histologic special types, particularly in the metastatic setting, are scarce. These studies have shown that although

the repertoire of somatic genetic alterations found in metastatic BCs is remarkably similar to that of primary tumors, *TP53*, *ESR1*, *ARID1A*, *ERBB2*, *GATA3*, *KMT2C*, *NCOR1*, *NF1*, and *RB*1 have been found to be significantly more frequently mutated in metastatic disease [58,63,66]. In addition, ER+ metastatic BCs have been shown to more frequently display the APOBEC mutagenesis and homologous recombination DNA repair deficiency (HRD) processes than primary ER+ disease [63,66]. Massively parallel sequencing studies by our group and others have revealed that some histologic special types of BC are underpinned by highly recurrent or even pathognomonic genetic alterations, including *ETV6-NTRK3* fusion gene in secretory carcinoma, and *MYB/MYBL1* rearrangements or *MYB* amplification in adenoid cystic carcinoma [60,67-68]. Furthermore, other primary special types of BC, albeit not driven by pathognomonic fusion genes or somatic mutations, have been found to harbor repertoires of genetic alterations that differ from those of primary IDC-NST [25,70-77]. In addition to *CDH1* mutations, primary invasive lobular carcinomas (ILCs) have been shown to display an enrichment in mutations affecting *PIK3CA*, *PTEN*, *TBX3*, *FOXA1*, *AKT1*, *ARID1A*, *ERBB2*, and *ERBB3*, primary mucinous BCs harbor a lower frequency 1q gains, 16q losses, and *PIK3CA* and *TP53* mutations than ER+ HER2- IDC-NSTs matched by clinical characteristics [70,78], micropapillary BCs display a repertoire of genetic alterations comparable to that of common forms of BCs, with frequent mutations in *PIK3CA*, *TP53*, *GATA3*, and *MAP2K4* [79], and metaplastic BCs, compared to triple-negative IDC-NSTs, more frequently harbor mutations affecting genes of the PI3K/AKT/mTOR and canonical Wnt pathways [77,80]. Here, through the reanalysis of targeted sequencing data generated with an FDA-approved multigene sequencing assay, we sought to define the repertoire of somatic genetic alterations of metastatic ILCs, mixed mucinous, micropapillary, and metaplastic BCs, and determine whether the landscape of somatic mutations and CNAs of metastatic special types

of BC is distinct from that of their primary counterparts or of metastatic IDC-NSTs.

3.2 Methods

3.2.1 Cases and study population

The study was approved by Memorial Sloan Kettering Cancer Center Institutional Review Board as part of the project whose findings were initially published by Razavi et al. [58]. Informed consent was provided in the original study by Razavi et al. [58]. Targeted massively parallel sequencing data of primary and metastatic BCs were obtained from the study by Razavi et al. [58]. All cases had been previously subjected to targeted capture massively parallel sequencing using the MSK-IMPACT sequencing assay from the study by Razavi et al. [58]. Following the criteria put forward by the WHO [1], 309 BCs were classified as of one of the special histologic types included in this study: 259 were classified as classic ILCs (n = 127 metastatic and n = 132 primary), 19 as mixed (i.e., >50% but <90% mucinous component) mucinous carcinomas (n = 5 metastatic and n = 14 primary), 20 as pure micropapillary carcinomas (n = 12 metastatic and n = 8 primary), and 11 as metaplastic BCs (n = 6 metastatic and n = 5 primary). The initial diagnosis of a given special histologic type of BC was retrieved from Razavi et al. [58], and cases for which the histologic material of the sample subjected to sequencing was available (n = 265) were reviewed centrally by a board-certified breast pathologist for diagnosis confirmation. Pleomorphic ILCs (metastatic, n = 6; primary, n = 8) were excluded from further analyses. ER and HER2 status had been assessed by immunohistochemistry and/or FISH, as previously described, following the American Society of Clinical Oncology/College of American Pathologists guidelines [20,98].

3.2.2 Comparison with common forms of breast cancer

For the comparison of non-synonymous TMB, FGA, frequency of non-synonymous somatic mutations, and CNAs, metastatic BCs of special histologic subtype were compared to those of IDC-NSTs included in the same study [58], matched by age (20-year intervals), menopausal status, and ER/HER2 status and to those of their primary counterparts. Metastatic ILCs were matched to metastatic IDC-NSTs from the study by Razavi et al. [58] previously subjected to MSK-IMPACT™at a 1:2 ratio, whereas mixed mucinous BCs, micropapillary BCs, and metaplastic BCs were matched to IDC-NSTs at a 1:3 ratio. No statistically significant differences were observed in the therapy received prior to tumor sampling between the metastatic BCs of special histologic types and metastatic IDC-NSTs matched by clinicopathologic characteristics in the cohorts analyzed in this study. Lollipop plots were produced using MutationMapper on cBioPortal, manually curated and mutation types were color-coded as follows: splice-site SNV (yellow), missense SNV (green), truncating SNV (black), in-frame insertion/ deletion (brown), and hotspot mutation (orange).

3.2.3 Targeted massively parallel sequencing analysis

All samples included in this study were subjected to targeted sequencing using the FDA-approved MSK-IMPACT™assay, as part of the study by Razavi et al. [58]. Non-synonymous somatic mutations, amplifications, and homozygous deletions were retrieved from the original study. The raw MSK- IMPACT™sequencing data (FASTQ files) were reprocessed using our validated bioinformatics pipeline, as previously described [99,100], for the inference of copy number gains and losses, and loss of heterozygosity of genes targeted by somatic mutations and mutational signatures.

Mutations affecting hotspot codons were annotated as described. Non-synonymous TMB was calculated as the number of non-synonymous mutations divided by the total genomic region assessed by MSK-IMPACT™, per megabase. The FGA, defined as the number of base pairs which are not copy neutral divided by the size of genome assayed, was retrieved from the original study by Razavi et al. [58]. Mutational signatures were defined using SigMA [87] using all synonymous and non-synonymous somatic mutations of cases with at least five SNVs. Tumor purity was inferred using FACETS [102]. The median tumor purity of special histologic type BCs analyzed in study was 0.43 (95% CI=0.30-0.87). Of note, the tumor purity of metastatic BCs of special histologic type (median = 0.47; 95% CI = 0.30-0.88) was higher than that of primary tumors (median = 0.39; 95% CI = 0.27-0.86; P = 1.4 × 10-2). As expected, the tumor purity of metastatic ILCs (median = 0.48; 95% CI = 0.30-0.88) was higher than that of primary ILCs (median = 0.38; 95% CI = 0.28-0.86; P = 1.1 × 10-3), whereas no differences were observed in the comparisons between metastatic and primary BCs of other histologic types analyzed in this study.

3.2.4 Assessment of TILs infiltration

Histologic assessment of TILs infiltration in primary and metastatic ILCs with a sufficient number of SNVs ($\geq$5) for accurate assessment of mutational signatures by SigMA [87], and available H&E slides was performed. The assessment of TILs infiltration was conducted following the guidelines described by the International TIL working group [103]. In brief, following the examination of one representative section, the intra-tumoral stromal area covered by mononuclear cells, including lymphocytes and plasma cells, was recorded.

3.2.5 Immunohistochemistry

Immunohistochemical analyses for MLH1, MSH2, MSH6, and PMS2 were performed in a Bond-3 automated stainer platform (Leica Biosystems, Wetzlar, Germany). In brief, following antigen retrieval (ER2, Leica) for 30-40min, tissue sections were incubated with monoclonal antibodies against MLH1 (clone ES05; Leica Biosystems; dilution 1:500), MSH2 (clone G219-1129; Cell Marque, Rocklin, CA; dilution 1:750), MSH6 (clone EP49; Dako, Glostrup, Denmark; dilution 1:500), or PMS2 (clone A16.4; BD Biosciences, Franklin Lakes, NJ; dilution 1:500) for 30 min. A polymer-based kit was employed as secondary reagent (Leica Biosystems). Assessment of the MLH2, MSH2, MSH6, and PMS2 expression was conducted by a board-certified pathologist following the current standard practice.

3.2.6 Statistical analysis

Statistical analyses were conducted using R v3.1.2. Fisher's exact tests were employed for comparisons between categorical variables, and Mann-Whitney U test were used for continuous variables. All tests were two-sided and P values < 0.05 were considered statistically significant. We performed multiple testing correction using the Benjamini–Hochberg procedure to control for the false discovery rate (q values). To assess the mutual exclusivity between *ERBB2* and *ESR1* mutations (hotspot mutations and non-hotspot pathogenic mutations) in ER+ metastatic ILC and IDC-NST using CoMEt [104]. The data generated and analyzed during this study are described in the following data record [105].

3.3 Results

3.3.1 Clinicopathologic characteristics

We reanalyzed the sequencing data corresponding to 309 samples of histologic special types of BC reported by Razavi et al. [58], comprising 154 and 155 primary and metastatic BCs, respectively. A total of 127 primary and 132 metastatic ILCs, 14 primary and five metastatic mixed mucinous BCs, 8 primary and 12 metastatic micropapillary BCs, and 5 primary and 6 metastatic metaplastic BCs were included in this study. Most primary and metastatic ILCs (95 and 81%), mixed mucinous BCs (79 and 100%), and micropapillary BCs (63 and 75%) were ER+ HER2-, whereas 80 and 83% of primary and metastatic metaplastic BCs were of triple-negative phenotype, respectively. We observed an enrichment of HER2-positive (7%) and ER+ HER2- (12%) phenotypes in metastatic ILCs, as compared to primary ILCs (2%, each), whereas primary ILCs were more frequently of ER+ HER2- (95%) phenotype than metastatic ILCs (81%; P = 1.6 × 10-3).

3.3.2 Repertoire of somatic genetic alterations in primary and metastatic ILCs

The ILCs included in this study were of classical type, characterized by a uniform population of small to medium-sized tumors cells with a dyshesive growth pattern, usually arranged in strands and single files (Fig. 3.1a). We compared the repertoire of somatic genetic alterations between primary and metastatic ILCs and observed that the non-synonymous tumor mutation burden (TMB) of metastatic ILCs (median = 4.2, 95% CI = 0.8-20.5) was significantly higher than that of primary ILCs (median 2.5, 95% CI = 0.8-8.2, P = 3.9 × 10-7, Mann–Whitney U test; Fig. 3.1b). The genes most frequently altered in metastatic ILCs (n = 132) overlapped with those reported in primary tumors, including *CDH1* (76%), *PIK3CA*

(52%), *TP53* (20%), *ERBB2* (19%), *FGF19*, *CCND1*, *FGF3*, *FGF4* (each, 17%), and *TBX3* (16%). Other frequently altered genes in metastatic ILCs included *ARID1A* and *FOXA1* (11%, each), *MAP3K1* (10%), and *PTEN* (9%; Fig. 3.1c). As compared to primary ILCs (n = 127), metastatic ILCs (n = 132) more frequently harbored genetic alterations affecting *TP53* (20% vs 9%, respectively; P = 1.3 × 10-2), *ESR1* (15% vs 2%, respectively; P=3×10-4), *FAT1* (9% vs 2%, respectively; P = 1.1 × 10-2), *RFWD2* (8% vs 1%, respectively; P = 5.4×10-3), and *NF1* (8% vs 2%, respectively; P=1.9×10-2; Fig. 3.1c). We also observed that *ERBB2* was numerically more frequently altered in metastatic ILCs than in primary ILCs (19% vs 12%, respectively; P=1.2×10-1; Fig. 3.1c): 12% (16/132) of metastatic ILCs harbored *ERBB2* mutations, 5% (6/132) *ERBB2* gene amplification, and 2% (3/132) harbored both *ERBB2* mutations and gene amplification. In 10% (13/132) of cases the *ERBB2* mutations were hotspot mutations in the kinase domain (Fig. 3.1c). Next, we compared the mutational repertoire of metastatic ILCs (n = 132) to that of metastatic IDC-NSTs matched by age, menopausal status, and ER/HER2 status at a 1:2 ratio (n=264). The nonsynonymous TMB of metastatic ILCs (median = 4.2, 95% CI = 0.8-20.5) was significantly higher than that of metastatic IDC-NSTs matched by clinicopathologic characteristics (median = 3.3, 95% CI = 0.8-12.8; P = 2.4 × 10-6; Mann–Whitney U test; Fig. 3.1b). Compared to age, menopausal, and ER/HER2 status-matched metastatic IDC-NSTs (n = 264), metastatic ILCs (n = 132) harbored a significantly higher frequency of genetic alterations affecting CDH1 (76% vs 3%, respectively; P = 9.8 × 10-54),*PIK3CA* (52% vs 34%, respectively; P = 7 × 10-4), *ERBB2* (19% vs 11%, respectively; P=2.8×10-2), *TBX3* (16% vs 6%, respectively; P=3.6×10-3), *NCOR* (9% vs 3%, respectively; P = 1 × 10-2), *RFWD2* (8% vs 1%, respectively; P = 6 × 10-4), and a significantly lower frequency of genetic alterations affecting *TP53* (20% vs 34%, respectively; P = 3.3 × 10-3), *ESR1* (15% vs 25%, respectively; P = 2.8 × 10-2), and *GATA3* (7% vs 20%, respectively; P = 4 × 10-4)

among others (Fig. 3.1c). Most *ERBB2* mutations identified in metastatic ILCs (n = 59%), primary ILCs (67%), and metastatic IDC-NSTs (45%) affected hotspot loci (Fig. 3.1d). Notably, the L755S *ERBB2* hotspot mutation was the most frequent in both metastatic (8/22; 36%) and primary ILCs (4/12; 33%; Fig. 3.1d), as previously reported [25]. This mutation, however, accounted for only 17% (2/12) of the *ERBB2* mutations detected in metastatic IDC-NSTs matched by clinicopathologic characteristics (Fig. 3.1d). Of note, we did not identify differences in pre-biopsy therapy of patients with metastatic ILCs harboring L755S *ERBB2* mutations, that could account for the observed enrichment. Given the role of *ESR1* mutations and *ERBB2* mutations in endocrine therapy resistance in ER+ metastatic BCs [34,58,81-84], we sought to investigate their mutual exclusivity in ER+ metastatic ILCs (n=113) and IDC-NSTs (n=226). We observed that hotspot mutations or pathogenic mutations affecting *ESR1* and *ERBB2* were mutually exclusive in metastatic ILCs (P = 4.8 × 10-2; CoMEt; Fig. 3.1e). These findings are consistent with those reported by Razavi et al. [58], where ESR1 and *ERBB2* mutations were found to be mutually exclusive in ER+ HER2- BCs regardless of their histologic subtype. Hence, akin to common cancer types of BC, *ESR1*, and *ERBB2* mutations are present in a mutually exclusive manner in metastatic ILCs, and may constitute mechanisms of resistance to endocrine therapy [34,58,81-84]. To define the repertoire of somatic genetic alterations present in ILCs, we combined the primary and metastatic ILCs of this study in one cohort (n = 259), and compared them to combined primary and metastatic IDC-NSTs, matched to the ILCs according to age, menopausal status, ER/HER2 status, and sample type at a 2:1 ratio (n = 518). This analysis revealed differences consistent with our findings when primary and metastatic ILCs were compared to IDC- NST separately. Combined primary and metastatic ILCs (n = 259) displayed a higher non-synonymous TMB (P = 1.8 × 10-7) than combined primary and

metastatic IDC-NSTs (n = 518). In addition, as compared to combined primary and metastatic IDC-NSTs, combined primary and metastatic ILCs harbored a higher frequency in genetic alterations affecting *CDH1* (79% vs 3%; P=1.0×10-116), *PIK3CA* (54% vs 37%; P= 5.3 × 10-6), *ERBB2* (15% vs 8%, P = 1.6 × 10-3), *TBX3* (13% vs 6%; P = 1.4 × 10-3), *ARID1A* (10% vs 5%; P = 3.4 × 10-2), *NCOR1* (8% vs 4%, P=4×10-2), *RUNX1* (7% vs 3%: P=2.7×10-2), and *RFWD2* (5% vs 1%, P=7×10-4), and a lower frequency of genetic alterations affecting TP53 (14% vs 33%; P = 2.6 × 10-8), *ESR1* (9% vs 15%, P=2.2×10-2), and *GATA3* (5% vs 19%; P=1.2×10-8).

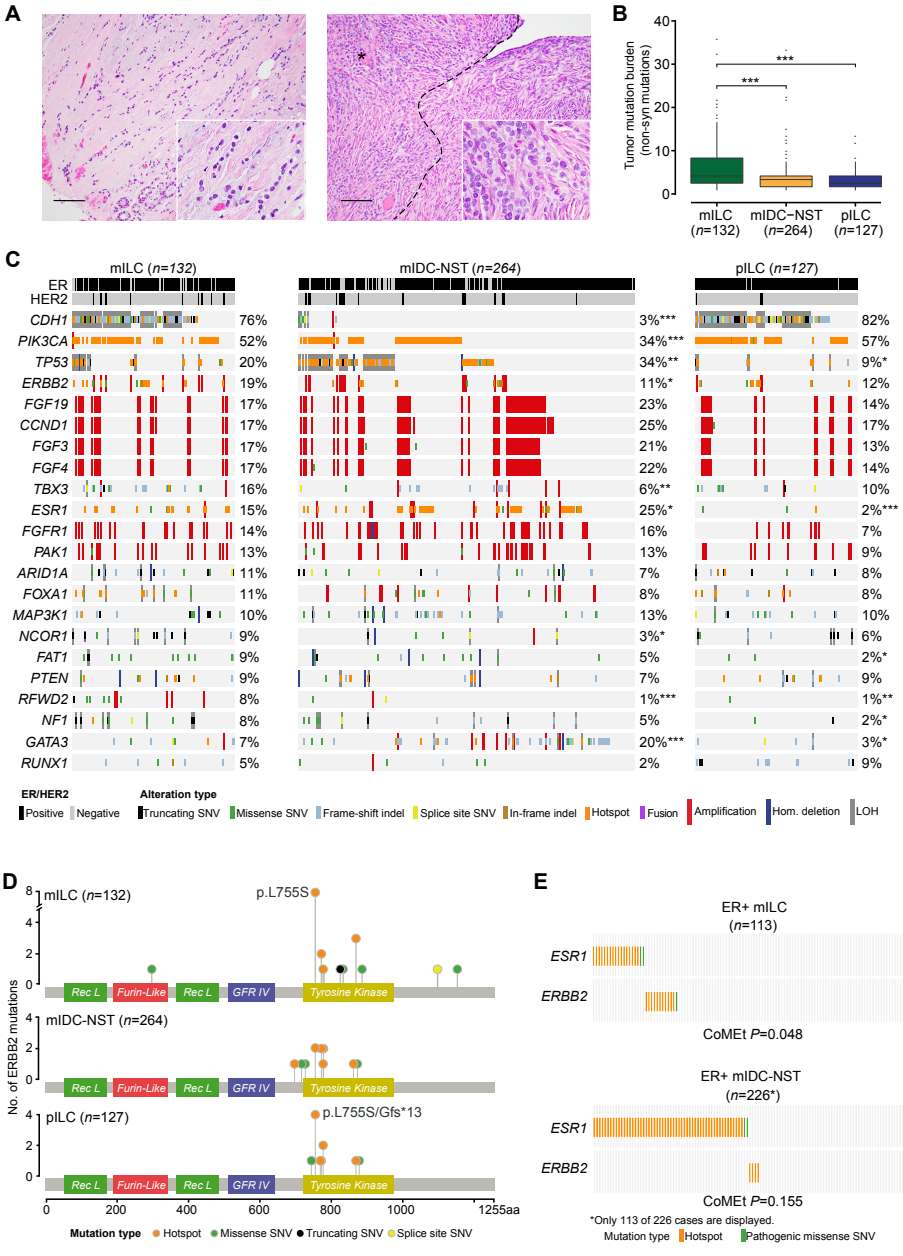

A
B
Tumor mutation burden (non-syn mutations)
mILC (n=132)
mIDC−NST (n=264)
pILC (n=127)
C
mILC (n=132)
mIDC-NST (n=264)
pILC (n=127)
ER
HER2
CDH1 76% 3%*** 82%
PIK3CA 52% 34%*** 57%
TP53 20% 34%** 9%*
ERBB2 19% 11%* 12%
FGF19 17% 23% 14%
CCND1 17% 25% 17%
FGF3 17% 21% 13%
FGF4 17% 22% 14%
TBX3 16% 6%** 10%
ESR1 15% 25%* 2%***
FGFR1 14% 16% 7%
PAK1 13% 13% 9%
ARID1A 11% 7% 8%
FOXA1 11% 8% 8%
MAP3K1 10% 13% 10%
NCOR1 9% 3%* 6%
FAT1 9% 5% 2%*
PTEN 9% 7% 9%
RFWD2 8% 1%*** 1%**
NF1 8% 5% 2%*
GATA3 7% 20%*** 3%*
RUNX1 5% 2% 9%
ER/HER2
Positive
Negative
Alteration type
Truncating SNV
Missense SNV
Frame-shift indel
Splice site SNV
In-frame indel
Hotspot
Fusion
Amplification
Hom. deletion
LOH
D
mILC (n=132)
p.L755S
mIDC-NST (n=264)
pILC (n=127)
p.L755S/Gfs*13
No. of ERBB2 mutations
Rec L
Furin-Like
Rec L
GFR IV
Tyrosine Kinase
0
200
400
600
800
1000
1255aa
Mutation type
Hotspot
Missense SNV
Truncating SNV
Splice site SNV
E
ER+ mILC (n=113)
ESR1
ERBB2
CoMEt P=0.048
ER+ mIDC-NST (n=226*)
ESR1
ERBB2
CoMEt P=0.155
*Only 113 of 226 cases are displayed.
Mutation type
Hotspot
Pathogenic missense SNV

FIGURE 3.1 *(previous page)*: ***Repertoire of genetic alterations in primary and metastatic invasive lobular carcinomas of the breast***. *A) Representative photomicrographs of a H&E-stained primary breast invasive lobular carcinoma (pILC; left) and a metastatic lobular carcinoma (mILC) involving ovarian stroma (right). Scale bars, 50 µm. B) Boxplots depicting the non-synonymous tumor mutation burden of mILCs (n = 132), metastatic invasive ductal carcinomas of no special type matched by age, menopausal status, and estrogen receptor (ER)/HER2 status (mIDC-NSTs; n = 264), and pILCs (n = 127). Mann–Whitney U test, two-tailed. C) Comparison of the cancer genes most frequently affected by non-synonymous somatic mutations, amplifications, or homozygous deletions in mILCs (n = 132), metastatic age-, menopausal status-, and ER/HER2 status- matched mIDC-NSTs (n = 264) and pILCs (n = 127). Cases are shown in columns and genes in rows. Mutation types are color-coded according to the legend. ER/HER2 status are shown on phenobars (top). *P $<$ 0.05, **P $<$ 0.01, ***P $<$ 0.001; Fisher's exact test, two-tailed. D) Schematic representation of the protein domains of ERBB2 and the somatic mutations in metastatic mILCs (n=132), mIDC-NSTs matched by clinicopathologic characteristics (n = 264) and pILCs (n = 127). Mutations are color-coded according to the legend, and their frequency is represented by the height of each lollipop (y-axis). E) Mutual exclusivity analysis of ESR1 and ERBB2 hotspot, and oncogenic/likely oncogenic mutations in ER+ mILCs (n = 113) and mIDC-NSTs (n = 226). Hom. homozygous, Indel insertion/deletion, LOH loss of heterozygosity, SNV single nucleotide variant.*

3.3.3 Repertoire of somatic genetic alterations in primary and metastatic mixed mucinous BCs

The mixed mucinous BCs analyzed in this study were characterized by areas of tumor cells floating in lakes of mucin admixed with areas of IDC-NST (Fig. 3.2a, b). Metastatic mixed mucinous BCs harbored a significantly higher non-synonymous TMB (median = 4.2, 95% CI = 1.7-5.8) than primary mixed mucinous BCs (median=0.8, 95% CI=0.8-2.2, P=1.5×10-3, Mann–Whitney U test), but comparable to that of metastatic IDC-NSTs matched according to clinical features (median = 3.3, 95% CI = 0.8-6.4, P =6×10-1; Fig. 3.2b). The repertoire of genetic alterations of metastatic

mixed mucinous BCs (n = 5) in this study was similar to that of primary mucinous/mixed BCs [70,71,85]. Although based on a small number of cases, this analysis revealed that the genes recurrently altered in metastatic mixed mucinous BC and not altered in primary mixed mucinous BCs of this study, and pure/mixed mucinous BCs reported by our group and others [70,71,85] included ESR1 (60% vs 0%, respectively; P = 1 × 10-2) and NCOR (40% vs 0%, respectively; P = 6 × 10-2; Fig. 3.2c). In agreement with previous studies [86], compared to metastatic IDC-NSTs matched by clinical features, metastatic mixed mucinous BCs harbored a higher frequency of 11q13.3 amplification (60% vs 7%, respectively; P = 3.2 × 10-2; Fig. 3.2c).

3.3.4 Repertoire of somatic genetic alterations in primary and metastatic micropapillary BCs

The micropapillary BCs included in this study were characterized by morula-like clusters of tumor cells without a fibrovascular core within pseudovascular spaces (Fig. 3.3a). We observed no significant differences in the non-synonymous TMB of metastatic micropapillary BCs (median = 1.2, 95% CI = 0.8-4) compared to that of primary micropapillary BCs (median = 1.2, 95% CI = 0.8-4; P = 1.3 × 10-1) or to that IDC-NSTs matched by clinical features (median = 3.3, 95% CI = 0.8-11.7; P = 3.9 × 10-1). In a way akin to IDC-NSTs, the most frequently altered genes in metastatic and primary micropapillary BCs were *PIK3CA* (58 and 25%) and *TP53* (42 and 38%). Recurrent alterations in ESR1 (25%), *KDR*, *ARID1B*, and *ATR* (17%, each) were restricted to metastatic micropapillary BCs (Fig. 3.3b). On the other hand, *MYC* gene amplification was more frequent in primary than in metastatic micropapillary BCs (38% vs 0%, respectively; P = 4.9 × 10-2; Fig. 3.3b).

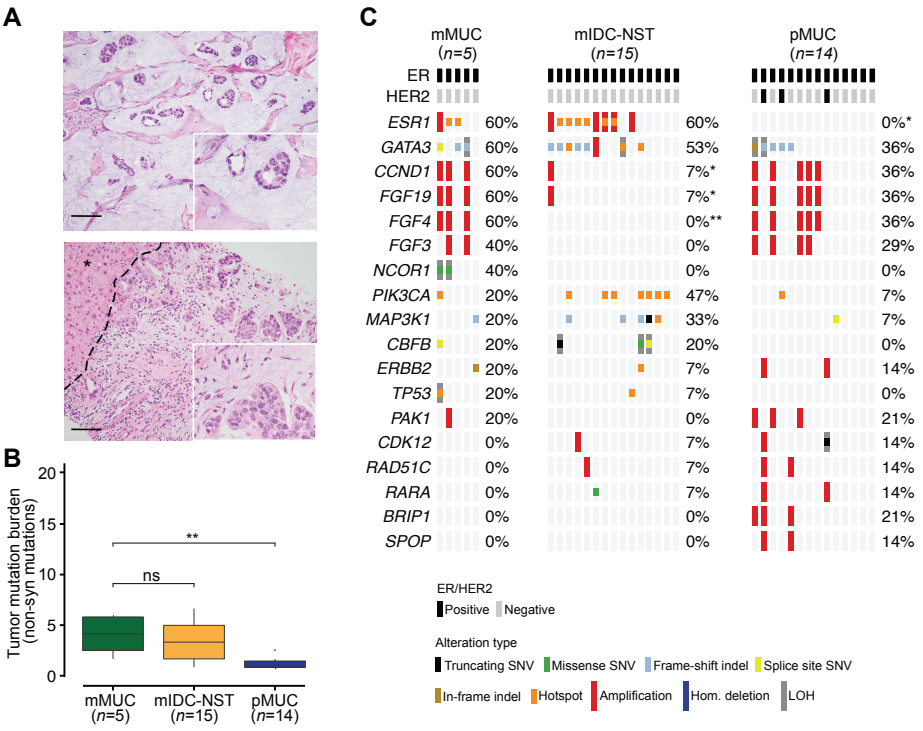
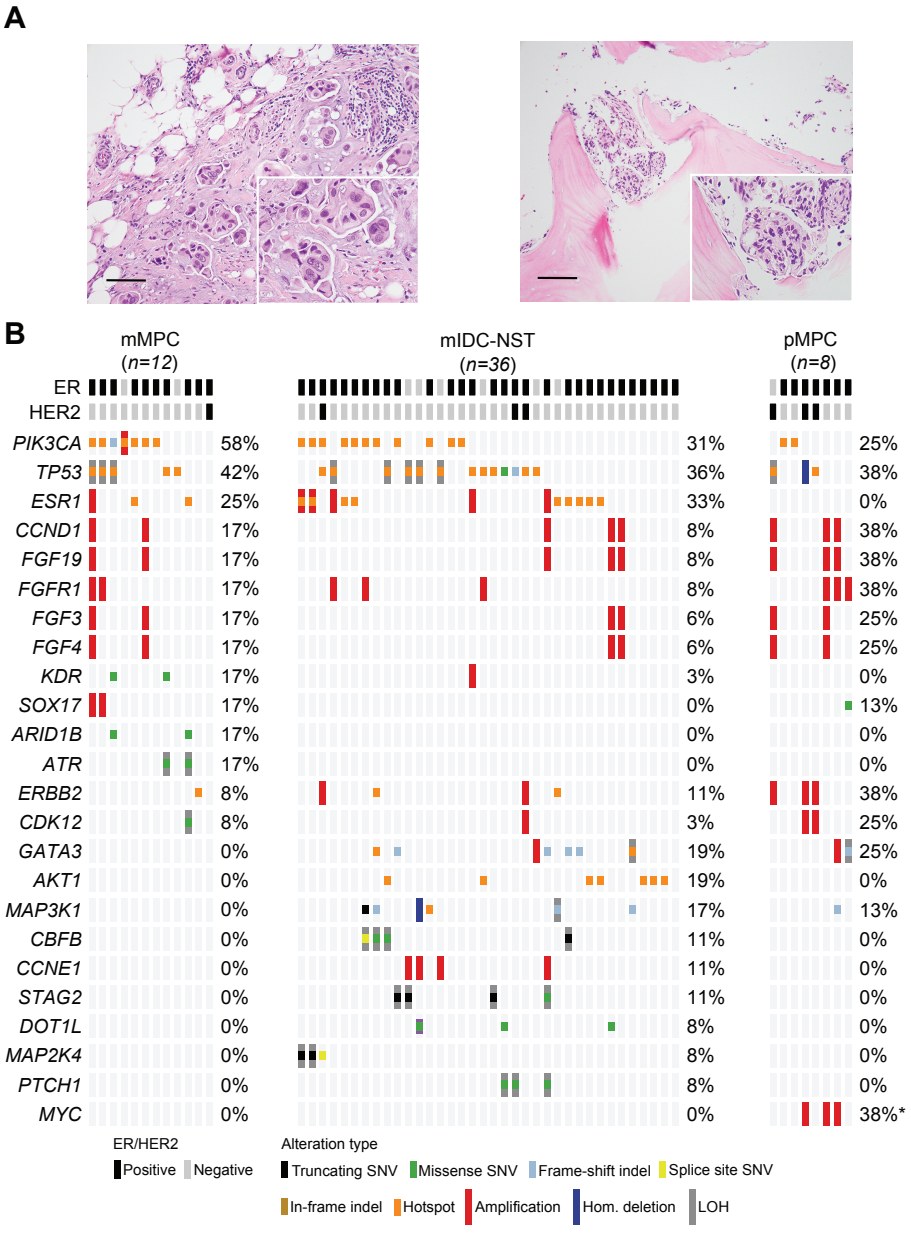


FIGURE 3.2: ***Repertoire of genetic alterations in primary and metastatic mixed mucinous breast cancers***. *A) Representative photomicrographs of a H&E-stained primary mixed mucinous breast cancer (pMUC; top), and a metastatic mixed mucinous breast cancer (mMUC) involving liver (bottom). Scale bars in a, 100 µm (top) and 50 µm (bottom). B) Boxplots depicting the non-synonymous tumor mutation burden in mMUCs (n = 5), metastatic invasive ductal carcinoma of no special type matched by age, menopausal status, and estrogen receptor (ER)/HER2 status (mIDC-NST; n = 15), and pMUCs (n = 14). Mann–Whitney U test, two-tailed. C) Comparison of the cancer genes most frequently affected by non- synonymous somatic mutations, amplifications, or homozygous deletions in mMUCs (n=5), in age-, menopausal status-, and ER/HER2 receptor status-matched mIDC-NSTs (n = 15), and in pMUCs (n = 14). Cases are shown in columns and genes in rows. Mutation types are color- coded according to the legend. ER/HER2 status are shown on phenobars (top). *P < 0.05; Fisher's exact test, two-tailed. Indel insertion/deletion, LOH loss of heterozygosity, SNV single nucleotide variant.*


A
B
mMPC
(n=12)
mIDC-NST
(n=36)
pMPC
(n=8)
ER
HER2
PIK3CA 58% 31% 25%
TP53 42% 36% 38%
ESR1 25% 33% 0%
CCND1 17% 8% 38%
FGF19 17% 8% 38%
FGFR1 17% 8% 38%
FGF3 17% 6% 25%
FGF4 17% 6% 25%
KDR 17% 3% 0%
SOX17 17% 0% 13%
ARID1B 17% 0% 0%
ATR 17% 0% 0%
ERBB2 8% 11% 38%
CDK12 8% 3% 25%
GATA3 0% 19% 25%
AKT1 0% 19% 0%
MAP3K1 0% 17% 13%
CBFB 0% 11% 0%
CCNE1 0% 11% 0%
STAG2 0% 11% 0%
DOT1L 0% 8% 0%
MAP2K4 0% 8% 0%
PTCH1 0% 8% 0%
MYC 0% 0% 38%*
ER/HER2
Positive
Negative
Alteration type
Truncating SNV
Missense SNV
Frame-shift indel
Splice site SNV
In-frame indel
Hotspot
Amplification
Hom. deletion
LOH